What Difference Does It Make? A Comparison of Health State Preferences Elicited From the General Population and From People with Multiple Sclerosis

Running title: What Difference Does It Make?

Elizabeth Goodwin PhD[1], Colin Green PhD[1,2], Annie Hawton PhD[1,2]

1 Health Economics Group, Institute of Health Research, College of Medicine and Health, University of Exeter, Exeter, UK

2 South West Collaboration for Leadership in Applied Health Research and Care (CLAHRC), University of Exeter Medical School, College of Medicine and Health, University of Exeter, Exeter, UK

Corresponding author: Elizabeth Goodwin, Health Economics Group, South Cloisters, St Luke's Campus, University of Exeter, Exeter UK EX1 2LU.
e.goodwin@exeter.ac.uk
+44 1392726073

Funding Statement: Financial support for this study was provided in part by the Multiple Sclerosis Society of Great Britain and Northern Ireland. Partial funding was received from the UK NIHR Collaboration for Leadership in Applied Health Research and Care of the South West Peninsula (PenCLAHRC) to CG and AH. The funding agreements ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. The views expressed in this publication are those of the authors and not necessarily those of the Multiple Sclerosis Society, the UK NIHR or the Department of Health.

Précis: The differences between public and patient valuations of multiple sclerosis health states are significant and complex, with important implications for the results of cost-effectiveness analyses.

Acknowledgements: Funding for this research was provided by the MS Society of Great Britain and Northern Ireland. The data for this paper was collected by the UK MS Register project (ref: 16/SW/0194) and Accent Marketing Research Ltd. We would like to acknowledge the contribution of all members of the UK MS Register team based at Swansea University Medical School, and all members of the team at Accent.

What difference does it make? A comparison of health state preferences elicited from the general population and from people with multiple sclerosis.

Abstract

Objectives: A major debate in the Quality-Adjusted Life-Year (QALY) literature concerns whose preferences should be used to estimate health state values (HSVs) and to calculate QALYs. This study explores differences between public and patient values for multiple sclerosis (MS) health states, described using an MS-specific classification system (MSIS-8D).

Methods: The MSIS-8D is an existing preference-based measure of health-related quality of life in MS, which has two tariffs of HSVs, based on the preferences of a representative sample of the UK general population (n=1702) and of people with MS living in the UK (n=1635), elicited using the time trade-off technique. Here we explore differences between HSVs by sample type, using descriptive statistics and multivariate regression methods.

Results: Overall, the survey of people with MS produced significantly higher HSVs; estimated values ranged from 0.079-0.883 for the general population survey, and 0.138-0.894 for the MS survey. Differences in HSVs were more pronounced for severe health states. The difference between patient and public values varied across the dimensions of the MSIS-8D. People with MS placed greater importance on *Cognition* than the general population, leading to lower HSVs when impairment was at a worse level; the reverse was true for the *Daily Activities*, *Fatigue* and *Depression* dimensions.

Conclusions: We identified significant differences in HSVs by sample type. Using patient rather than public values may influence the results of economic evaluations, depending upon the dimensions of health-related quality of life affected by the intervention being assessed, and may therefore have important consequences for reimbursement decisions.

Highlights

What is already known about the topic?

Previous research has shown that preferences elicited from members of the general population frequently differ from those of patients for the same health states. However, comparisons between population and patient preferences have typically relied upon small samples of respondents and health states, limiting their ability to examine these differences in detail.

What does the paper add to existing knowledge?

This paper compares two full tariffs of health state values for the same condition-specific preference-based measure (the MSIS-8D): one based on the preferences of a representative sample of the UK general population and one from a representative sample of people with MS.

Analysis provides a detailed examination of the differences between these two groups in terms of the relative importance of individual dimensions of health-related quality of life and the impact of moving between levels on each dimension.

What insights does the paper provide for informing health care-related decision making?

Insights are provided into how the choice of sample for a preference elicitation survey may affect the results of cost-effectiveness analyses, and ultimately influence resource allocation decisions.

Introduction

Decisions on the funding of healthcare interventions are frequently informed by evidence concerning their cost-effectiveness. Cost-effectiveness is commonly expressed as the cost-per-QALY, or quality-adjusted life-year, of the intervention [1]. QALYs assess both length and quality of life as a single measure, with length of life adjusted by a weight to reflect the quality of life during this time. These quality of life (QALY) weights, or health state values (HSVs), are usually estimated from preferences given from samples of the general population, but they have also been estimated from samples of people who have the condition that the treatment is designed to address ('patients') [2, 3].

In publicly funded healthcare systems, it is often argued that societal preferences should guide resource allocation in order to reflect the views of those who are funding the service. Conversely, the theory of welfare economics posits that the well-being of a society equals the sum of the utilities of its individual members, implying that it is more appropriate to base decisions regarding public programmes on the preferences of those set to gain or lose directly from the decision, ie patients. Furthermore, patients are likely to have more experience of poor health and may be considered better placed to value how this affects quality of life [3]. These considerations have led to suggestions that public preferences may be more suitable for system-wide decision-making, while patient values should be used to inform condition-specific resource allocation and individual-level treatment decisions [4].

A number of studies have compared preferences elicited from patients with preferences elicited from the general population for the same health states. Significant differences have been found, with most studies reporting that patients provide higher HSVs compared to the

general population [5-12]. A higher HSV implies that the health state is more preferred, and is associated with better health-related quality of life (HRQL). Several reasons have been put forward to explain this phenomenon. For example it has been suggested that patients provide higher values as a result of *scale recalibration* [13], which occurs when a person's interpretation of a scale alters in line with changes in their own internal reference point. This may occur due to experience of ill health, aging, changes in social comparators or other changes in circumstances [3]. Another possible reason is the ability of patients to contextualise health state descriptions. The brief, standardised nature of these descriptions may be insufficient to fully convey the experience of inhabiting a health state, and patients may be more capable of filling in the gaps in these descriptions on the basis of their knowledge or experience [7, 14]. Frequently, the additional dimensions of well-being that patients consider when valuing a health state are not adversely affected by the disease [15]. Conversely, members of the public are likely to experience *framing effects*, in which the framing of health state scenarios around specific dimensions highlights the presence of decrements in those dimensions in the absence of a wider context [16]. Another potential cause is that healthy people may underestimate the extent to which they would adapt to negative changes in their health. Patients are more likely to have experienced adaptation to poor health states, and will take this into account in their preferences [17].

Studies focusing on mental health [18-21] or dementia [22] health states, however, have found that patients provided lower HSVs compared to the general population, and some research indicates that the relative importance of particular dimensions of HRQL may differ [12, 18, 23]. Furthermore, differences between patient and public values can vary according to the severity of the health state, with patients providing lower values for the mildest health states and higher values for the most severe states [6, 19-22, 24], resulting in "valuation

compression", ie a narrower range of HSVs [25]. . Taken together, these differences can have sufficient impact on the results of cost-effectiveness studies to influence resource allocation decisions, and the direction and nature of their effects can vary [3]. Despite the salience of the issue, the debate over whether public or patient preferences should be used to value health states has yet to be resolved [26].

The scope of previous studies has been somewhat limited, comparing directly observed valuations of a subset of health states, elicited from relatively small samples of patients and the general population. In previous research, the Multiple Sclerosis Impact Scale – Eight Dimensions (MSIS-8D) was developed. The MSIS-8D is a preference-based measure for multiple sclerosis (MS), which has two sets of HSVs that can be used to inform cost-effectiveness analysis: one based on the preferences of the general population [27] and one on the preferences of people with MS [28]. MS is a long-term neurological condition, which frequently follows a relapsing-remitting course but can be progressive from outset, and causes a wide and variable range of physical, psychological and cognitive symptoms [29]. Little is known about how the preferences of the general public may differ from those of people with MS.

This research aims to compare the two full tariffs of values that were estimated for the MSIS-8D and to consider what implications any differences between these HSVs may have for resource allocation decisions. On the basis of the literature outlined above, we hypothesise that people with MS will value health states more highly, place greater weight on non-physical dimensions of HRQL, and provide a narrower range of values across all health states.

Methods

Data

The data used in this study were from two previous valuation surveys, one undertaken with a sample of the general population and one with a sample of people with MS [27, 28]. . The same methods were used for both surveys, enabling comparisons to be drawn [30]. The surveys employed the Measurement and Valuation of Health (MVH) version of the time trade-off (TTO) technique to elicit HSVs [31] and were administered via the internet to maximise sample size and representativeness [32]. The Supplementary Material provides additional detail on sample selection and recruitment, procedures for health states considered worse than being dead, issues regarding online administration of the TTO and steps taken to enhance data quality. The TTO technique asks respondents to choose between Life B, which offers a particular health state for ten years, or Life A, which offers full health for a shorter period of time (x). In these questions HSVs are calculated as $x/10$. The MVH process also allows respondents to assign negative values to health states they consider worse than being dead [33]. Health states were described using the MSIS-8D classification system, which represents eight dimensions of importance to the HRQL of people with MS: physical functioning, mobility, social activities, daily activities, fatigue, cognitive function, emotional well-being and depression [34]. Each dimension has four levels. Preferences were elicited for a sample of 169 MSIS-8D health states, selected using the Rasch vignette approach [2] to reflect states that are likely to be experienced by people with MS at different levels of severity. The health states were stratified into five groups according to severity and selected at random to produce blocks of five health states. Each respondent valued one block of five

health states, plus the worst MSIS-8D state. Respondents completed the MSIS-8D descriptive system for their own health over the previous two weeks, and a total raw score was calculated as a measure of their current health status. Data were collected on age, gender, presence of dependent children in the household, and socio-economic status. The general population sample was recruited through a market research panel, using quota sampling to ensure a representative sample of the UK population. People with MS were recruited from the UK MS Register, which has been found to be representative of people with MS in the UK in terms of characteristics including gender, age at onset and MS type [35].

Analysis methods

Analysis of the extent to which the two sets of HSVs differed was undertaken in two phases. All analysis was undertaken in Stata version 14.

Phase 1

The methods used in Phase 1 are similar to those reported by Mulhern et al [36]. Data from both surveys were combined into a single dataset. Mean HSVs from the two valuation surveys were compared using a two-sample, two-sided *t*-test. To explore whether differences between patient and public values varied according to the severity of the health state [6, 19, 20, 24, 37], separate *t*-tests were carried out for each of five severity groups (constructed previously in order to select balanced sets of states for respondents to value), plus the worst MSIS-8D state.

Regression analysis was undertaken, based on the standard model that defines the value of a health state as follows [2]:

$h_{ij} = f(\beta' X_{\lambda\partial}) + \varepsilon_{ij}$

i: individual health states

j: individual respondents

h_{ij}*:* value for health state *i* valued by respondent *j*;

X: vector of dummy variables for each level λ of dimension ∂, with level λ = 1 as baseline;

β: coefficients on X

ε_{ij}: error term.

We adopted a random effects specification, as this was found to be the best performing of the models previously considered for estimating HSVs for the MSIS-8D [27, 28]. Two models were estimated:

- Model 1 includes a dummy variable for sample-type (patient or public).
- Model 2 includes the sample-type variable and significant socio-demographic variables.

Model 1 aimed to determine whether the sample-type (general population or people with MS) was a significant predictor of HSVs, independently of the levels on the dimensions of the MSIS-8D. Model 2 aimed to determine whether the sample-type variable remained significant when controlling for differences in socio-demographic characteristics and respondents' own MSIS-8D scores. All variables that were found to be significantly associated with HSVs in bivariate analysis were included in this model.

Further regression analyses using Model 1 included the addition of interaction terms between each dimension of the MSIS-8D and the sample-type dummy variable, enabling us to explore whether differences between patients and the public were statistically significant at an individual dimension level, using F-tests.

Phase 2

In the second phase, we compared the published algorithms, recommended for use in generating HSVs for MSIS-8D health states from the perspective of the UK general population [27] and people with MS [28]. In each case, one preferred model had been selected from a number of candidate models, based on predictive ability. Candidate models included individual-level and mean-level ordinary least squares, random effects and random effects Tobit specifications [38, 39]. In these preferred models, where dimension-levels were represented by non-significant coefficients, the affected levels were merged to produce additional versions of the models [40]. The predictive ability of the models was assessed in terms of the mean absolute error and the proportion of health states with errors greater than 0.05 or 0.10 [2]. Here, an "error" is defined as the difference between the observed and the predicted value for a health state, hence smaller numbers are preferred.

The sizes of the coefficients were compared to investigate differences in the weighting of individual dimensions and in the effect of moving between levels on each dimension, and the ranges of HSVs predicted by the models were compared.

Results

Comparisons of descriptive data

Data from 1576 members of the general population and 1596 people with MS were included in the analysis. Responses from 126 members of the general population and 39 people with MS were excluded for providing inconsistent or illogical responses. The median number of observations per health state was 47 for the general population survey and 48 for the survey of people with MS. Analysis of data quality is provided in the Supplementary Material.

Table 1 reports socio-demographic data for both samples. Differences between the samples reflected known differences between people with MS and the general population in the UK: a higher percentage of female respondents [29], a lower proportion of respondents aged 45 years or under [41], and a lower proportion of respondents in employment [42] in the MS sample.

The mean observed values for each health state (Appendix 1) ranged from 0.08 for the worst health state to 0.89 for the best when valued by the general population, and from 0.15 to 0.94 when valued by people with MS.

Phase 1 analysis

The results of the *t*-tests (Table 2) show a significant difference in the mean HSVs derived from the two groups. The overall mean HSV was 0.052 higher when derived from the preferences of people with MS. The difference in means was significant ($p<0.01$) for all severity groups.

The results of Models 1 and 2 are summarised in Table 3. Model 1 reflects the expected pattern of preferences: as dimension-levels increase, reflecting a greater negative impact on health status, HSVs decrease. The coefficient for 'having MS' (a variable that distinguishes between the public and patient samples) is significant, with a positive effect, indicating that people with MS reported higher HSVs for MSIS-8D states. Model 2 shows that older age and having dependent children were significant predictors of higher HSVs, while being in a "severe" MSIS-8D health state was associated with lower HSVs. Inclusion of these variables reduced the size of the coefficient for 'having MS' from 0.052 ($p<0.001$) to 0.042 ($p=0.002$). As there was no significant difference between the samples in the proportion of respondents with dependent children, this indicates that part of the difference in HSVs is explained by the differences in the age profile of the two samples, and that this is partly offset by the differences in health status (Table 1).

Analysis of the interactions between sample type and dimensions of the MSIS-8D found that the differences between patients and the public were not statistically significant at an individual dimension level ($p>0.05$).

Phase 2 analysis

Table 4 compares the preferred models recommended for estimating MSIS-8D values from the perspective of the UK general population (the "Public Model") and of people with MS (the "MS Model"). The results of these models are represented graphically in Figure 2.

HSVs based on public preferences ranged from 0.079 to 0.882, while those based on the preferences of people with MS ranged from 0.138 to 0.893. The MS Model had a slightly smaller mean absolute error and fewer health states with prediction errors greater than 0.1. When health states were valued by people with MS, each increase in problems with *Cognition* (represented by moving from level 1 to the level corresponding to each model coefficient) had a greater negative impact on HSVs than when health states were valued by the general population. This resulted in larger coefficients for each level of the *Cognition* dimension in the MS Model.

Conversely, when health states were valued by members of the general population, each increase in problems with *Daily Activities*, *Fatigue* and *Depression* had a greater negative impact on HSVs than when valued by people with MS, resulting in larger coefficients for all levels of these dimensions in the Public Model.

For the remaining dimensions, the differences between patients and the public varied between levels, with a mix of larger and smaller coefficients.

Table 4 and Figure 2 also illustrate the impact on HSVs of moving between adjacent levels on each dimension, ie from level 1 to 2, from level 2 to 3, and from level 3 to 4. Here, the differences between people with MS and the general population are represented by the differences in the distance between adjacent model coefficients. This reveals a more complex picture, in which most dimensions had a mix of smaller and larger single-level increments when the two models are compared.

Discussion

Discussion of the results

We have found clear differences in HSVs derived using preferences elicited from people with MS compared to those of the general public. As hypothesised, the analysis of observed HSVs indicates that people with MS valued health states more highly than the general population, ie they were prepared to give up less time in full health in order to avoid health problems associated with MS. This difference was significant across the full range of severity, although it was less pronounced in the mildest health states, suggesting that differences between patient and public values vary according to the severity of the health state as anticipated. A similar pattern emerged from the regression analysis: the type of sample (general population or people with MS) was a significant predictor of HSVs, and the range of predicted values was narrower when based on the preferences of people with MS (0.755 compared to 0.803 for the general population), providing some support for the hypothesis that patient values are subject to valuation compression.

The methods used here provide a robust basis for the comparison of two preference elicitation surveys for the same classification system, using large, representative samples of the UK general population and of people with MS. Earlier studies have reported similar findings [5-8, 10, 12, 19-22, 24]. However, as far as we are aware, this is the first study to estimate full tariffs from both perspectives for the same classification system, using a choice-based technique. This has enabled us to undertake more detailed analysis of the impact of eliciting preferences from these two groups across the full range of dimensions (and levels) of HRQL.

A key advantage of this approach is that it provides the opportunity to investigate whether people with MS and members of the general population apply different relative weights to

individual dimensions of HRQL. On the basis of previous research [19-23], we hypothesised that people with MS would give greater weight to dimensions associated with mental and cognitive health. This hypothesis was only partially supported. Contrary to expectations, the model based on the preferences of the general population had larger coefficients for the *Depression, Fatigue and Daily Activities* dimensions than the model based on the preferences of people with MS, indicating that the general population placed greater weight on these dimensions. Coefficient sizes were larger when based on the preferences of people with MS for the *Cognition* dimension, reflecting the findings of Rowen and colleagues [22]. Although the Phase 1 analysis found no statistically significant differences between the coefficients, the size and direction of the differences that were apparent in the Phase 2 analysis may influence the apparent cost-effectiveness of interventions that affect these dimensions. For example, interventions targeting the impact of MS on cognitive function may appear more cost-effective if assessed using HSVs based on the preferences of people with MS [23].

For the purposes of economic evaluation, the changes in HSVs that occur when moving between health states are more relevant than their absolute size. If all health states were valued 0.02 higher by patients, both versions of the model would produce identical values for any given change in health status (excluding any impacts from mortality). It is the difference in the relative weighting of dimensions and in the effect of changes in individual dimension-levels that is relevant in determining the effects on the results of cost effectiveness analysis. For example, a shift from health state 32432222 (level 4 on the *Mobility* dimension, level 3 on *Physical* and *Daily Activities*, level 2 on all other dimensions) to health state 22322222 (level 3 on *Mobility*, level 2 on all other dimensions) produces an improvement in HSV of 0.05 (0.71 – 0.66) using the model based on public preferences, compared to 0.11 (0.80 – 0.69) using the model based on the preferences of people with MS. Conversely, a shift from

health state 22224222 (level 4 on *Fatigue*, level 2 on all other dimensions) to health state 22223221 (level 3 on *Fatigue*, level 1 on *Depression*, level 2 on all other dimensions) produces an improvement of 0.09 (0.74 – 0.65) using the model based on public preferences, compared to 0.04 (0.78 – 0.74) using the model based on the preferences of people with MS. Discrepancies of this magnitude in the value of a shift from one health state to another may have serious implications for the results of economic evaluations, and could influence resource allocation decisions [3].

Limitations of study

The main limitation of this study is that it is quantitative in nature and therefore cannot explain *why* preferences for health states differ between the general population and people with MS. We are currently exploring this, using cognitive interviews. This will help to inform the debate regarding whether preferences should be elicited from patients or from the general public when estimating QALY weights for a condition-specific PBM [38].

In order to achieve large, representative samples of the general population and people with MS living across the UK, valuation surveys were administered online. Previous studies have shown that the quality of TTO data can suffer in the absence of an interviewer [32, 43-45]. Therefore, steps were taken to maximise data quality (see supplementary material).

Part of the difference in HSVs was explained by the differing age profiles of the samples: overall, the sample of people with MS was older than the general population sample. In both surveys, older people valued health states more highly, ie they were less willing to give up life-years to avoid suboptimal health states, than younger respondents. This difference in age

profile reflects an acknowledged difference between the two populations[41], hence the results still represent a real difference between people with MS and the general population. It is, however, possible that the difference in HSVs between age groups was an artefact of the valuation technique employed, in which respondents were asked to trade life-years against HRQL [12]. This could be investigated by exploring alternative preference elicitation techniques, such as the standard gamble, discrete choice experiments, or alternative versions of the TTO, using different time horizons.

Possible reasons for differences in values

One of the interesting features of the results is that respondents whose own current health status was classified as “severe” provided significantly lower HSVs than other respondents, independently of whether they had MS. Given that respondents with MS were more likely to be in a “severe” health state (Table 1), this may shed some light on the reasons why patient and public values differ. It does not appear to be the differences in the current health status of the two groups of respondents that causes people with MS to value health states more highly than the general population. Instead, this finding suggests that the causal mechanisms underpinning the differences between public and patient values are perhaps more aligned to attitudinal differences associated with the experience, and the process, of living with a chronic condition such as MS. This is an important finding, which merits further investigation beyond the scope of this paper.

Of the suggested mechanisms by which patients value health states more highly, few offer a convincing explanation for why the difference between public and patient values is reversed for certain dimensions of HRQL. A possible cause relates to differences between individual

domains of HRQL in terms of patients' capacity to adapt to decrements. As Dolan and Kahneman [17] have suggested, some health states may be easier to adapt to than others, providing a normative argument to support the use of patient preferences to inform resource allocation decisions. However, patient preferences themselves are not immune to distortions, and are prone to value compression, which can reduce the apparent cost-effectiveness of interventions. Consequently, Dolan and Kahneman have proposed novel methods for valuing health states according to experienced utility, rather than hypothetical preferences. Meanwhile, within the current methodological framework, the debate over "whose values" should inform resource allocation continues. We recommend using both MSIS-8D and MSIS-8D-P values within sensitivity analyses, alongside the preferred PBM for a given jurisdiction (eg the EQ-5D in the UK), to provide broader, contextual information about the likely cost-effectiveness of treatments for MS, for use in resource allocation decision-making [4, 26-28, 46, 47].

Conclusion

We have shown that, overall, eliciting preferences from people with MS generates higher values for MSIS-8D health states than HSVs based on the preferences of the general public. Importantly, however, the impact of using patient values on the results of cost-effectiveness analyses will vary, depending upon the specific dimensions of HRQL that are affected by the intervention being assessed. The choice of "whose preferences" to use could therefore have important consequences for reimbursement decisions.

References

1. NICE. Guide to the methods of technology appraisal 2013. National Institute for Health and Care Excellence (NICE), 2013. http://www.nice.org.uk/article/pmg9/chapter/foreword. Accessed 23rd October 2017.

2. Brazier JE, Rowen D, Mavranezouli I, et al. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). Health Technol Assess. 2012;16:1-114.

3. Brazier J, Akehurst R, Brennan A, et al. Should patients have a greater role in valuing health states? Appl Health Econ Health Policy. 2005;4:201–8.

4. Gold MR, Siegel JE, Russell LB, Weinstein, MC. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996.

5. Boyd NF, Sutherland HJ, Heasman KZ, Tritchler DL, Cummings BJ. Whose values for decision making? Med Decis Making. 1990;10:58–67.

6. De Wit GA, Busschbach JJ, De Charro FT. Sensitivity and perspective in the valuation of health status: whose values count? Health Econ. 2000;9:109–26.

7. Happich M, von Lengerke T. Valuing the health state 'tinnitus': Differences between patients and the general public. Hear Res. 2005;207:50–5.

8. Lenert LA, Treadwell JR, Schwartz CE. Associations between health status measures and utilities: implications for policy. Med Care. 1999;37:479-89.

9. Llewellyn-Thomas H, Sutherland HJ, Tibshirani R, Ciampi A, Till JE, Boyd NF. The measurement of patients' values in medicine. Med Decis Making. 1982;2:449-62.

10. Peeters Y, Stiggelbout AM. Health state valuations of patients and the general public analytically compared: A meta-analytical comparison of patient and population hHealth state utilities. Value Health. 2010;13:306-9.

11. Sackett DL, Torrance GW. The utility of different health states as perceived by the general public. J Chronic Dis. 1978;31:697–704.

12. Krabbe PFM, Tromp N, Ruers TJM, van Riel PLCM. Are patients' judgments of health status really different from the general population? Health Qual Life Outcomes. 2011;9:1-9.

13. Ubel PA, Peeters Y, Smith D. Abandoning the language of "response shift": a plea for conceptual clarity in distinguishing scale recalibration from true changes in quality of life. Qual Life Res. 2010;19:465-71.

14. Jansen SJ, Stiggelbout AM, Wakker PP, Nooji MA, Noordijk EM, Kievit J. Unstable preferences: a shift in valuation or an effect of the elicitation procedure? Med Decis Making. 2000;20:62-71.

15. Stiggelbout AM, de Vogel-Voort E. Health state utilities: a framework for studying the gap between the imagined and the real. Value Health. 2008;11:76-87.

16. Ubel P, Loewenstein G, Jepson C. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. Qual Life Res. 2003;12:599-607.

17. Dolan P, Kahneman D. Interpretations of utility and their implications for the valuation of health. Econ J. 2008;118:215-34.

18. Brazier J. Measuring and valuing mental health for use in economic evaluation. J Health Serv Res Policy. 2008;13:70–5.

19. Gerhards SA, Evers SM, Sabel PW, Huibers MJ. Discrepancy in rating health-related quality of life of depression between patient and general population. Qual Life Res. 2011;20:273-9.

20. Papageorgiou K, Vermeulen K, Schroevers MJ, et al. Do individuals with and without depression value depression differently? And if so, why? Qual Life Res. 2015;24:2565-75.

21. Pyne JM, Fortney JC, Tripathi S, Feeny D, Ubel P, Brazier J. How bad is depression? Preference score estimates from depressed patients and the general population. Health Serv Res. 2009;44:1406-23.

22. Rowen D, Mulhern B, Banerjee S, et al. Comparison of general population, patient, and carer utility values for dementia health states. Med Decis Making. 2015;35:68-80.

23. Rand-Hendriksen K, Augestad LA, Kristiansen IS, Stavem K. Comparison of hypothetical and experienced EQ-5D valuations: relative weights of the five dimensions. Qual Life Res. 2012;21:1005-12.

24. Mulhern B, Rowen D, Snape D, et al. Valuations of epilepsy-specific health states: a comparison of patients with epilepsy and the general population. Epilepsy Behav. 2014; 36:12-7.

25. Kind P, Dolan P. The effect of past and present illness experience on the valuations of health states. Med Care. 1995;33: AS255-AS263.

26. Menzel P, Dolan P, Richardson J, Olsen JA. The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis. Soc Sci Med. 2002;55:2149-58.

27. Goodwin E, Green C, Spencer A. Estimating a preference-based index for an eight dimensional health state classification system derived from the Multiple Sclerosis Impact Scale (MSIS-29). Value Health. 2015;18:1025-36.

28. Goodwin E, Green C, Hawton A. Health state values derived from people with multiple sclerosis for a condition-specific preference-based measure: Multiple Sclerosis Impact Scale - Eight Dimensions – Patient version (MSIS-8D-P). Value Health. 2018;21:1338-45.

29. Zajicek J, Freeman J, Porter B. Mulitple Sclerosis Care: A Practical Manual. Oxford: Oxford University Press, 2007.

30. Brazier J, Tsuchiya A. Preference-based condition-specific measures of health: what happens to cross programme comparability? Health Econ. 2010;19:125-9.
31. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997;35:1095-108.
32. Bansback N, Tsuchiya A, Brazier J, Anis A. Canadian valuation of EQ-5D health states: preliminary value set and considerations for future valuation studies. PLoS ONE. 2012;7(2):e31115. doi:10.1371/journal.pone.0031115..
33. Gudex C. Time trade-off user manual: Props and self-completion methods. York: The MVH Group, Centre for Health Economics, University of York, 1994.
34. Goodwin E, Green C. A QALY measure for multiple sclerosis: Developing a patient-reported health state classification system for an MS-specific preference-based measure. Value Health. 2015;18:1016-24.
35. Ford DV, Jones KH, Middleton RM, et al. The feasibility of collecting information from people with multiple sclerosis or the UK MS Register via a web portal: characterising a cohort of people with MS. BMC Med Inform Decis Mak. 2012;12:1-8.
36. Mulhern B, Rowen D, Brazier J, et al. Development of DEMQOL-U and DEMQOL-PROXY-U: generation of preference-based indices from DEMQOL and DEMQOL-PROXY for use in economic evaluation. Health Technol Assess. 2013;17:1-140.
37. Insinga RP, Fryback DG. Understanding differences between self-ratings and population ratings for health in the EuroQOL. Qual Life Res. 2003;12:611–19.
38. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. Measuring and valuing health for economic evaluation. Oxford: Oxford University Press, 2007.
39. Goodwin E, Green C. A systematic review of the literature on the development of condition-specific preference-based measures of health. Appl Health Econ Health Policy. 2016;14:161-83.

40. Versteegh MM, Leunis A, Uyl-de Groot CA, Stolk EA. Condition-specific preference-based measures: benefit or burden? Value Health. 2012;15:504-13.

41. Mackenzie IS, Morant SV, Bloomfield GA, MacDonald TM, O'Riordan J. Incidence and prevalence of multiple sclerosis in the UK 1990–2010: a descriptive study in the General Practice Research Database. J Neurol Neurosurg Psychiatry. 2014;85:76-84.

42. Employment that works: Supporting people with MS in the workplace. London: All-Party Parliamentary Group for MS, 2016. https://www.mssociety.org.uk/sites/default/files/Employment%20that%20works%20-%20APPG%20report.pdf. Accessed 15th February 2018.

43. Attema A, Edelaar-Peeters Y, Versteegh M, Stolk E. Time trade-off: one methodology, different methods. Eur J Health Econ. 2013;14:S53–S64..

44. Norman R, King M, Clarke D, Viney R, Cronin P, Street D. Does mode of administration matter? Comparison of online and face-to-face administration of a time trade-off task. Qual Life Res. 2010;19:499–508.

45. Versteegh M, Attema A, Oppe M, Devlin N, Stolk E. Time to tweak the TTO: results from a comparison of alternative specifications of the TTO. Eur J Health Econ. 2013;14:S43–S51..

46. Nord E, Pinto JL, Richardson J, et al. Incorporating societal concerns for fairness in numerical valuations of health programs. Health Econ. 1999 8:25–39.

47. Versteegh MM, Brouwer WBF. Patient and general public preferences for health states: A call to reconsider current guidelines. Soc Sci Med. 2016;165:66-74.

Table 1 Socio-demographic characteristics of the samples (general public and people with MS)

	General public (n=1576)		People with MS (n=1596)	
Gender				
Female	829	53%	1145	73%
Male	747	47%	424	27%
Age group				
16 to 25	206	13%	7	0%
26 to 35	269	17%	85	5%
36 to 45	261	17%	304	19%
46 to 55	279	18%	504	32%
56 to 65	253	16%	463	30%
Over 65	308	20%	205	13%
Dependent children				
No	1148	73%	1091	71%
Yes	428	27%	443	29%
Employment status				
Employed	801	51%	633	41%
Not employed	770	49%	912	59%
Socio-economic group				
A or B	422	26.78%	900	59.92%
C1 or C2	790	50.13%	403	26.83%
D or E	364	23.10%	199	13.25%
Own MSIS-8D score				

Mild	1118	70.94%	442	27.69%
Moderate	297	18.85%	686	42.98%
Severe	161	10.22%	468	29.32%
Diagnosed with MS	9	0.57%	1596	100%

Socio-economic group is defined according to the Approximated Social Grade Categories developed by the UK Office for National Statistics.

Own MSIS-8D scores are calculated as the sum of the raw scores over the eight items of the descriptive system.

Mild MSIS-8D_total score = 8 - 16

Moderate MSIS-8D_total score = 17 - 24

Severe MSIS-8D_total score = 25 - 32

nb Figures may not sum to the total number of respondents due to missing data for some socio-demographic variables.

Table 2 Results of t-tests to investigate differences between observed health state values from the general public and people with MS

		General public	People with MS	Difference	t stat	p-value
All health states	Mean HSV	0.484	0.536	-0.052	-7.878	<0.001
	SD	0.467	0.450			
Severity Group 1	Mean HSV	0.817	0.845	-0.027	-3.364	<0.001
(least severe)	SD	0.233	0.224			
Severity Group 2	Mean HSV	0.682	0.739	-0.057	-5.263	<0.001
	SD	0.319	0.289			
Severity Group 3	Mean HSV	0.578	0.639	-0.061	-4.695	<0.001
	SD	0.383	0.351			
Severity Group 4	Mean HSV	0.473	0.531	-0.058	-3.953	<0.001
	SD	0.421	0.398			
Severity Group 5	Mean HSV	0.269	0.318	-0.048	-2.849	0.004
	SD	0.483	0.474			
Worst health state	Mean HSV	0.083	0.146	-0.063	-3.699	<0.001
	SD	0.480	0.480			

Observations for all health states: public = 9,456; people with MS = 9,576; degrees of freedom = 19030

Observations for each severity group: public = 1576; people with MS =1596; degrees of freedom = 3170

HSV = health state value

SD = standard deviation

Table 3 Results of regression analysis for both samples combined, showing significance of having MS (Model 1) and of socio-demographic variables (Model 2)

		Model 1		Model 2	
Dimension	**Level**	**Coefficient**	**p-value**	**Coefficient**	**p-value**
Physical	2	-0.048	<0.001	-0.049	<0.001
	3	-0.052	0.001	-0.056	0.001
	4	-0.162	<0.001	-0.164	<0.001
Social	2	-0.013	0.264	-0.012	0.305
	3	-0.031	0.068	-0.032	0.065
	4	-0.073	<0.001	-0.077	<0.001
Mobility	2	-0.012	0.276	-0.012	0.289
	3	-0.027	0.096	-0.024	0.151
	4	-0.089	<0.001	-0.086	<0.001
Daily activities	2	-0.013	0.286	-0.011	0.371
	3	-0.027	0.106	-0.025	0.136
	4	-0.06	0.002	-0.058	0.004
Fatigue	2	-0.005	0.619	-0.005	0.648
	3	-0.017	0.272	-0.015	0.334
	4	-0.064	0.001	-0.060	0.002
Emotion	2	-0.016	0.122	-0.017	0.104
	3	-0.042	0.006	-0.042	0.007
	4	-0.078	<0.001	-0.081	<0.001
Cognition	2	-0.02	0.063	-0.018	0.096

	3	-0.043	0.006	-0.042	0.009
	4	-0.103	<0.001	-0.105	<0.001
Depression	2	-0.012	0.251	-0.016	0.146
	3	-0.055	<0.001	-0.057	<0.001
	4	-0.149	<0.001	-0.146	<0.001
Constant		0.861	<0.001	0.753	<0.001
Demographics					
Having MS		0.052	<0.001	0.042	0.002
Age group (years)	26 to 35			0.035	0.185
	36 to 45			0.068	0.007
	46 to 55			0.107	<0.001
	56 to 65			0.117	<0.001
	Over 65			0.103	<0.001
Dependent children	Yes			0.065	<0.001
Employed	No			0.015	0.254
Socio-economic group	C1 or C2			0.018	0.154
	D or E			0.019	0.231
MSIS-8D group	Moderate			-0.008	0.549
	Severe			-0.033	0.042
Performance					
Observations		19302		18102	
Groups		3217		3017	
Overall R^2		0.300		0.310	
Wald chi^2		15345.2		15988.34	
Wald chi^2 p-value		<0.001		<0.001	

Mild MSIS-8D_total score = 8 - 16

Moderate MSIS-8D_total score = 17 - 24

Severe MSIS-8D_total score = 25 - 32

Basecase for socio-demographic variables: age group 18-25; no dependent children, employed or self-employed; socio-economic group A or B; MSIS-8D group mild. Gender is not included in this analysis because it was not found to be associated with health state values in earlier bivariate analysis.

Table 4 Differences between the models used to produce tariffs for the MSIS-8D from the general public and from people with MS

Dimension	Model coefficients				Impact of moving between levels			
	Level	Public sample	MS sample	Diff	Level	Public sample	MS sample	Diff
Physical	2	-0.053**	-0.047**	-0.006	1 to 2	0.053	0.047	+0.006
	3	-0.060**	-0.065**	+0.006	2 to 3	0.006	0.018	-0.012
	4	-0.185**	-0.175**	-0.010	3 to 4	0.125	0.110	+0.015
Social	2	0‡	0‡	0.000	1 to 2	0.000	0.000	0.000
	3	-0.028*	-0.032*	+0.004	2 to 3	0.028	0.032	-0.004
	4	-0.079**	-0.067**	-0.012	3 to 4	0.051	0.036	+0.016
Mobility	2	-0.022	-0.003	-0.019	1 to 2	0.022	0.003	+0.019
	3	-0.022†	-0.003†	-0.019	2 to 3	0.000	0.000	0.000
	4	-0.069**	-0.077**	+0.008	3 to 4	0.047	0.074	-0.027
Daily activities	2	-0.024	0‡	-0.024	1 to 2	0.024	0.000	+0.024
	3	-0.024†	-0.020	-0.004	2 to 3	0.000	0.020	-0.020
	4	-0.064**	-0.048*	-0.016	3 to 4	0.040	0.028	+0.012
Fatigue	2	-0.026*	0‡	-0.026	1 to 2	0.026	0.000	+0.026
	3	-0.026†	-0.021	-0.005	2 to 3	0.000	0.021	-0.021
	4	-0.088**	-0.063**	-0.025	3 to 4	0.062	0.042	+0.020
Emotion	2	0‡	-0.015	+0.015	1 to 2	0.000	0.015	-0.015
	3	-0.041**	-0.042*	+0.001	2 to 3	0.041	0.026	+0.014
	4	-0.084**	-0.069**	-0.015	3 to 4	0.044	0.027	+0.016
Cognition	2	-0.014	-0.027*	+0.014	1 to 2	0.014	0.027	-0.014
	3	-0.014†	-0.052**	+0.038	2 to 3	0.000	0.025	-0.025

	4	-0.072**	-0.116**	+0.044	3 to 4	0.058	0.064	-0.006
Depression	2	-0.029*	0‡	-0.029	1 to 2	0.029	0.000	+0.029
	3	-0.074**	-0.040**	-0.034	2 to 3	0.045	0.040	+0.005
	4	-0.161**	-0.140**	-0.022	3 to 4	0.088	0.100	-0.012
Constant		0.882	0.893	-0.011				

Model performance		
Coefficients:	18	19
significant (p<0.05)	15 (83.33%)	15 (78.95%)
Overall R-squared	0.2951	0.3013
Wald chi^2 (p-value)	7950.80	8893.62
	(<0.001)	(<0.001)
Mean absolute error	0.0414	0.0364
Errors > 0.1	11 (6.51%)	3 (1.78%)
Errors > 0.05	52 (30.77%)	52 (30.77%)
Range of HSVs	0.079 to 0.882	0.138 to 0.893
Observations (groups)	9456 (1576)	9576 (1596)

‡levels 1 and 2 merged; †levels 2 and 3 merged; * p<0.05; **p<0.01

Diff = difference:

+ indicates that the size of the coefficient or the impact of moving between levels was greater when based on the preferences of people with MS

- indicates that the size of the coefficient or the impact of moving between levels was smaller when based on the preferences of people with MS

HSV = health state value